THE JACKKNIFE: STUDENT-T CONFIDENCE LIMITS AND SUBSAMPLING IN BLOCKS

by

D. V. Hinkley*

Technical Report No. 262

March 10, 1976

University of Minnesota
Minneapolis, Minnesota

*This research was supported by National Science Foundation Grant No. MPS75-08778.

SUMMARY

This article presents a simplified account of the theory of jackknife confidence limits with reference to the effects of s-at-a-time omission. Both the naive and a modified Student-t approximation are considered for the standardized jackknifed estimator. Monte Carlo results for the cases of variance and correlation parameters indicate the inference robustness of Student-t approximations and suggest the possible utility of multiple omission.

Key words: jackknife, variance, correlation, cross-validation

SUMMARY

This article presents a simplified account of the theory of jackknife confidence limits with reference to the effects of s-at-a-time omission. Both the naive and a modified Student-t approximation are considered for the standardized jackknifed estimator. Monte Carlo results for the cases of variance and correlation parameters indicate the inherent robustness of Student-t approximations and suggest the possible utility of multiple omission.

Key words: jackknife, variance, correlation, cross-validation

1. Introduction

Jackknifing of a sample statistic is useful in reducing bias and in permitting calculation of a robust standard error for the statistic. After Tukey's (1958) extension of Quenouille's (1956) proposal, most theoretical work on the jackknife method has focussed on large-sample properties in quite complex situations; see Miller (1974). The usual description of the jackknife method for homogeneous data involves one-at-a-time omission, since this is correctly thought to be more efficient than s-at-a-time omission for $s \geq 2$; theoretical work in support of one-at-a-time omission is reported by Rao and Webster (1966). However, with moderately large samples, and without the support of highspeed computing machinery, the s-at-a-time omission scheme may save a large amount of computing, while sacrificing little in statistical efficiency; this point is made by Mosteller and Tukey (1968, pp. 144-145). Of course in non-homogeneous situations, such as arise with stratified or hierarchical sampling, some blocking of the data may be necessary to obtain validity of the jackknife method; see Kish and Frankel (1974).

This article presents a simplistic account of jackknife theory including the effects of general s-at-a-time omission. Both the naive and a modified Student-t approximation are considered for the standardized jackknife estimator. Monte Carlo results for jackknife confidence intervals on variance and correlation parameters indicate the usefulness of Student-t approximations and suggest the possible utility of multiple omission.

The basic theory is outlined in Section 2. In Section 3 a data-based t-approximation is tentatively suggested. Simulation results for variance and correlation parameters are reported in Section 4, following which some tentative conclusions and additional remarks are made in Section 5.

1. Introduction

Il n'est guère douteux qu'un lien ait existé entre la méthode de Descartes et la géométrie algébrique naissante. Déjà dans une note jointe à la traduction latine de la Géométrie (1649), Frans van Schooten le jeune (1615?-1660) s'efforçait de mettre en relief l'originalité de la méthode de Descartes par rapport à celle de Viète (1591). Mais ce n'est que vers la fin du XIXe siècle que l'on a commencé à étudier systématiquement les manuscrits de Descartes et à publier sa correspondance. Ce travail a été mené à bien par C. Adam et P. Tannery dans leur édition des Œuvres de Descartes (1897-1913). Il a été complété par J. Vrin (1930-1974) qui a publié, entre autres, les Cogitationes privatae (1859), le Compendium musicae (1618), le Traité de mécanique (1634-1637) et les Olympica (1619). Ces différentes publications ont permis de mieux comprendre la genèse de la géométrie cartésienne. Elles ont montré que la méthode de Descartes n'était pas une création ex nihilo, mais qu'elle s'inscrivait dans une tradition mathématique bien précise, dont elle constituait l'aboutissement. Elles ont également montré que la méthode de Descartes était étroitement liée à sa physique et à sa métaphysique.

Il est aujourd'hui admis que la méthode de Descartes a été élaborée entre 1619 et 1628, date à laquelle Descartes a rédigé les Regulae ad directionem ingenii. Cependant, la publication de la Géométrie en 1637 a été précédée de la publication du Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences (1637), qui en constitue une sorte d'introduction. Descartes y expose les principes généraux de sa méthode, qu'il illustre ensuite dans la Géométrie. Il est donc nécessaire de tenir compte de ces différentes publications pour comprendre la méthode de Descartes.

Nous nous proposons ici d'étudier la méthode de Descartes telle qu'elle apparaît dans la Géométrie. Nous nous efforcerons de montrer que cette méthode est fondée sur une analyse des problèmes géométriques qui est inspirée de celle de Pappus. Nous montrerons ensuite que la méthode de Descartes repose sur une théorie des équations algébriques qui est également inspirée de celle de Viète. Nous montrerons enfin que la méthode de Descartes est étroitement liée à sa physique et à sa métaphysique.

2. Some Elementary Theory for Homogeneous Data

In the simplest situation to which the jackknife applies, an observation $y = (y_1,\ldots,y_n)$ is available on the random variable $Y = (Y_1,\ldots,Y_n)$, where these n components are exchangeable. Suppose that $T_n = t(Y)$ is an estimator for the scalar parameter μ, and let the vector Y be partitioned randomly into g groups or blocks each of size s; we assume that $n = gs$ for simplicity. The following is a brief description of the jackknife method for reducing bias of T_n and estimating its standard error in the above set-up.

Define

$$T_{n,s;-j} = t(Y \text{ omitting block } j) \qquad (j = 1,\ldots,g) \qquad (1)$$

and the pseudo-value

$$P_{n,s;-j} = gT_n - (g-1)T_{n,s;-j} \qquad (j = 1,\ldots,g). \qquad (2)$$

Then the jackknifed estimator of μ is

$$T^*_{n,s} = g^{-1} \sum_{j=1}^{g} P_{n,s;-j} \qquad (3)$$

and the estimated variance of $T^*_{n,s}$ (or T_n) is $V_{n,s}$ where

$$V_{n,s} = \{g(g-1)\}^{-1} \sum_{j=1}^{g} (P_{n,s;-j} - T^*_{n,s})^2 . \qquad (4)$$

The extent to which $V_{n,s}$ is an accurate estimate of $\text{var}(T^*_{n,s})$ depends on the accuracy with which T_n may be approximated by an unweighted average $n^{-1}\sum a(Y_j)$; when such a representation is exact, (4) is a sample variance. To understand the effects of different choices of s it is instructive to examine this special case in some detail.

Suppose, then, that

$$T_n = n^{-1} \sum_{j=1}^{n} a(Y_j) = n^{-1} \sum_{j=1}^{n} X_j , \qquad (5)$$

1. The Elementary Theory for Homogeneous Data

In the simplest situation to which the jackknife applies, an observation $y = (y_1,\dots,y_n)$ is available on the random variable $Y = (Y_1,\dots,Y_n)$, where these n components are exchangeable. Suppose that $T_n = t(Y)$ is an estimator for the scalar parameter μ, and let the vector Y be partitioned randomly into g groups or blocks each of size h; we assume that $n = gh$ for simplicity. The following is a brief description of the jackknife method for reducing bias of T_n and estimating its standard error in the above set-up.

Define

$$T_{n,g;-j} = t(Y \text{ omitting block } j) \qquad (j = 1,\dots,g) \tag{1}$$

and the pseudo-value

$$P_{n,g;j} = gT_n - (g-1)T_{n,g;-j} \qquad (j = 1,\dots,g). \tag{2}$$

Then the jackknifed estimator is

$$T^*_{n,g} = g^{-1}\sum_{j=1}^{g} P_{n,g;j} \tag{3}$$

and the estimated variance of $T^*_{n,g}$ (or T_n) is $\hat{V}_{n,g}$ where

$$\hat{V}_{n,g} = \{g(g-1)\}^{-1}\sum_{j=1}^{g}(P_{n,g;j} - T^*_{n,g})^2 . \tag{4}$$

The extent to which $\hat{V}_{n,g}$ is an accurate estimate of $V_{n,g} = \mathrm{var}(T^*_{n,g})$ depends on the accuracy with which T_n may be approximated by an unweighted average $n^{-1}\sum s(Y_j)$; when such a representation is exact, (4) is a sample variance. To understand the effects of different choices of g it is instructive to examine this special case in some detail.

Suppose, then, that

$$T_n = n^{-1}\sum_{j=1}^{n} s(Y_j) = n^{-1}\sum_{j=1}^{n} Z_j , \tag{5}$$

say; in practice the form $a(\cdot)$ may be unknown, of course. It is convenient to relabel $X_1,\ldots,X_n$ as $X_{11},\ldots,X_{1s}$, $X_{21},\ldots,$ $\ldots,X_{g1},$ $\ldots,X_{gs}$ to ease reference of particular blocks. Straightforward application of the definitions (1) - (4) then shows that

$$P_{n,s;-j} = \bar{X}_{j\cdot}, \quad T^*_{n,s} \equiv T_n = \bar{X}_{\cdot\cdot} \tag{6}$$

and

$$V_{n,s} = \{g(g-1)\}^{-1} \sum_{j=1}^{g} (\bar{X}_{j\cdot} - \bar{X}_{\cdot\cdot})^2 , \tag{7}$$

If we assume T_n to be an estimator of $\mu = E(X)$ and denote $var(X)$ by σ^2, then it follows that

$$E(T^*_{n,s}) \equiv \mu, \quad var(T^*_{n,s}) \equiv n^{-1}\sigma^2 , \tag{8}$$

$$E(V_{n,s}) \equiv n^{-1}\sigma^2 , \tag{9}$$

and

$$n^2 var(V_{n,s}) = \frac{2\sigma^4}{g-1} \left(1 + \frac{\gamma_2}{2s}\right) , \tag{10}$$

where γ_2 is the fourth standardized cumulant of X. The last result is an important one. If we were to use a Student-t approximation to

$$\frac{(T^*_{n,s} - \mu)}{\sqrt{V_{n,s}}} , \tag{11}$$

then the nominal degrees of freedom would be $g-1$, and if $\gamma_2 = 0$ the necessary chi-square approximation to $V_{n,s}$ would be very accurate. However, in general $\gamma_2 \neq 0$ and it is clear from (10) that the chi-square approximation to $V_{n,s}$ is more accurate the larger is s. Thus loss in

degrees of freedom may be accompanied by increased reliability of the Student-t approximation to (11) when s increases.

In general the form (5) only holds approximately, the full representation in regular problems being expressible as

$$T_n = \mu + b_n + n^{-1} \Sigma\, c(Y_j) + n^{-1}\, d_n(Y) \ , \qquad (12)$$

where $b_n = o(1)$ is a constant, $E\{c(Y)\} = 0$, and $d_n(Y) = o_p(1)$. For most common problems b_n may be expanded in powers of n^{-1}, and the term in n^{-1} is then absent from $T^*_{n,s}$. The relevance of the above discussion about $V_{n,s}$ and the Student-t approximation to (11) then depends on how negligible $d_n(Y)$ is, and on the accuracy of normal approximation to $n^{-1} \Sigma\, c(Y_j)$. As to the latter, this is controllable to some extent in that we may often work with a transformed version of T_n (such as the logarithm of a sample variance or the Fisher z-transform of a sample correlation coefficient) which is known to reduce the skewness and/or kurtosis of T_n.

The effect of the remainder term $d_n(Y)$ on $V_{n,s}$ may depend very much on s, and we might be concerned that the results (9) and (10), now only approximations, are inaccurate for usefully large values of s. Theoretical investigation of this problem is very complicated, and possibly a little misleading since large-sample approximations would inevitably be required. The alternative approach of Monte Carlo investigation is pursued on a small scale in Section 4 for two commonly-used statistics.

ода до установления безразличия по отношению ко всем возможным
вариантам выбора $\hat{\varphi}_{\alpha,\mu}$ в модели (4.1) от подстановки $\hat{\varphi}_{\alpha,\mu}$
вместо $\varphi_{\alpha,\mu}$.

Далее для простоты будем считать (4) моделью, удовлетворяющей
условиям регулярности, налагаемым на модель наименьших квадратов
в работе [5].

$$y_{\alpha,\mu} = \varphi_{\alpha,\mu}(\beta) + \varepsilon_{\alpha,\mu} \quad (4.2)$$

где $\varepsilon_{\alpha,\mu}$ – независимые случайные величины с нулевыми средними
и дисперсиями $\sigma^2_{\alpha,\mu}$, $\alpha = 1, \ldots, k$; $\mu = 1, \ldots, n_\alpha$.

Пусть $\hat{\beta}$ – оценка параметра β по методу наименьших квадратов,
$\hat{\varphi}_{\alpha,\mu} = \varphi_{\alpha,\mu}(\hat{\beta})$, $\hat{\varepsilon}_{\alpha,\mu} = y_{\alpha,\mu} - \hat{\varphi}_{\alpha,\mu}$ – остатки.

Обозначим $\hat{\sigma}^2_{\alpha,\mu}$ оценку дисперсии $\sigma^2_{\alpha,\mu}$, полученную по остаткам
$\hat{\varepsilon}_{\alpha,\mu}$ с помощью какой-либо квадратической формы $\hat{\sigma}^2_{\alpha,\mu} = \sum_{\alpha,\mu} \hat{\varepsilon}^2_{\alpha,\mu}$.

Рассмотрим вопрос о выборе наилучшей квадратической формы $\sum_{\alpha,\mu}$.
Для этого воспользуемся результатами работы [6], в которой
показано, что при выполнении условий регулярности, налагаемых
в [5], и при $n_\alpha \to \infty$ ($\alpha = 1, \ldots, k$) оценки $\hat{\beta}$ и $\hat{\sigma}^2_{\alpha,\mu}$ асимптотически
независимы, причем $\hat{\beta}$ распределена асимптотически нормально
с математическим ожиданием β и ковариационной матрицей
$\|\lambda^{-1}_{ij}\|$, где $\lambda_{ij} = \sum_{\alpha,\mu} \frac{1}{\sigma^2_{\alpha,\mu}} \frac{\partial \varphi_{\alpha,\mu}}{\partial \beta_i} \frac{\partial \varphi_{\alpha,\mu}}{\partial \beta_j}$.

Отсюда следует, что $\hat{\varphi}_{\alpha,\mu}$ также распределена асимптотически
нормально с математическим ожиданием $\varphi_{\alpha,\mu}$ и ковариационной
матрицей $\|\lambda_{ij}\|$, где $\lambda_{ij} = \sum_{\alpha,\mu} \frac{1}{\sigma^2_{\alpha,\mu}} \frac{\partial \varphi_{\alpha,\mu}}{\partial \beta_i} \frac{\partial \varphi_{\alpha,\mu}}{\partial \beta_j}$.

Таким образом, $\hat{\varphi}_{\alpha,\mu}$ можно рассматривать как оценку
параметра $\varphi_{\alpha,\mu}$ по методу наименьших квадратов, а $\hat{\varepsilon}_{\alpha,\mu}$ – как
остатки, удовлетворяющие условиям регулярности, налагаемым
в работе [5] на остатки модели (4.1).

Поэтому для $\hat{\varepsilon}_{\alpha,\mu}$ справедливы все результаты работы [5],
в частности, теорема 1, согласно которой для того, чтобы
$\sum_{\alpha,\mu} \hat{\varepsilon}^2_{\alpha,\mu}$ была несмещенной оценкой $\sum_{\alpha,\mu} \sigma^2_{\alpha,\mu}$, необходимо и достаточно,
чтобы $\sum_{\alpha,\mu} = \sum_{\alpha,\mu} \lambda_{\alpha,\mu} \lambda_{\alpha,\mu}$, где $\lambda_{\alpha,\mu}$ – произвольные числа.

Отсюда следует, что $\sum_{\alpha,\mu} = \sum_{\alpha,\mu} \lambda_{\alpha,\mu} \lambda_{\alpha,\mu}$ – единственная
несмещенная оценка $\sum_{\alpha,\mu} \sigma^2_{\alpha,\mu}$, которая может быть получена
из $\hat{\sigma}^2_{\alpha,\mu}$ с помощью квадратической формы $\sum_{\alpha,\mu}$.

Таким образом, $\hat{\sigma}^2_{\alpha,\mu}$ является единственной несмещенной
оценкой $\sigma^2_{\alpha,\mu}$, которая может быть получена из $\hat{\varepsilon}_{\alpha,\mu}$ с помощью
квадратической формы $\sum_{\alpha,\mu}$.

Отсюда следует, что $\hat{\sigma}^2_{\alpha,\mu}$ является единственной несмещенной
оценкой $\sigma^2_{\alpha,\mu}$, которая может быть получена из $\hat{\varepsilon}_{\alpha,\mu}$ с помощью
квадратической формы $\sum_{\alpha,\mu}$.

3. A Simple Modification to the t-approximation

It was suggested in the discussion following equation (10) that for small s the jackknife variance estimate $V_{n,s}$ could behave quite non-normally due to the fourth cumulant γ_2. Thus the nominal $g-1$ degrees of freedom associated with $V_{n,s}$ might lead to inaccurate Student-t approximations. This is indeed the case when T_n is $\log_e$(sample variance) or $\tanh^{-1}$(sample correlation) and $s = 1$. We therefore consider a simple modification to the degrees of freedom.

The basic approach we take is to approximate the distribution of $nV_{n,s}$ by that of $\sigma^2\chi_d^2 \div d$; to estimate d from the data; and to use d as a replacement for $g-1$ as the degrees of freedom in the Student-t approximation for (11). There are two fairly extreme choices for the estimation of d, one of which is to use (10) together with the raw moment estimate c_2 of γ_2 to obtain

$$(g-1)^{-1}\{1 + c_2/(2s)\} = d^{-1} ;$$

the other extreme is to jackknife $nV_{n,s}$ in order to obtain an estimate of its variance, and equate that estimate to $2n^2V_{n,s} \div d$, which is the estimated variance of $\sigma^2\chi_d^2 \div d$. The latter approach is more in the spirit of the jackknife, but clearly would lead to complicated calculation in practice. We opt for the following compromise. Consider $gV_{n,s}$ as defined by (4) to be the sample variance of g numbers $P_{n,s;-j}$ (rather than a complicated function of the original data), and use the formula for jackknife variance of a sample variance to estimate the variance of $gV_{n,s}$. This leads simply to

.6 _____

mil dada (6l) acisarge mitvellal mibsenvilo eda nl bodsaqiyn nav ll

sdarp ovoled lmoo $\hat{\mathbf{v}}_{s,a}$ rebmirbna ouaulsyr ellmirbal edl a .lmuM

sontgdb l-g .lsuluom ads ambl $\hat{\mathbf{v}}_{s,a}$ mesldmms ldtuvl edd ol sub tilmrmov-nom

u-dmsbnvd odlmpovnd ol bsol dilgms $\hat{\mathbf{v}}_{s,a}$ dsbr bsdsoosvs.mbsvsdlo

rosvlisv olgnmv)$_{\mathbf{g}}$lodl sl $\hat{\mathbf{u}}_{\mathbf{g}}$ mobm sesp nlr bssbdl nl sldl .vssbnlmlmoxqqs

silqvlp s tsbfsmos svolovnd ovl .l r v .bsb (mobslmtmoo slgmvd)$^{\wedge}$_dmvd`lvd

.mubsvrl lo tsolbob sdd ol mobsbsldlldom

lo mobvludsrsl .ldl ndsvnlxqlpv ol sl sdsd sv dosonqlpn olsed sld

b .nrlv ol bln .sdsd sdd movd .b sgmvdsbs ol .i $\hat{\mathbf{u}}_{\mathbf{g}}$ $\hat{\mathbf{v}}_{s,a}$ lo vmnls vp

u-smobvl.s sdd .l mobsbd lo sssvpsb srld sb .l-n .vsd dmovrslqsn s sn

sdd nol movnlvd rus.dvs vlnlvnlowd sve nsd'l .(ll) vsd sobsvmlmovqls

dsspdvn vsd srld ddlv rsolsqvol (ll) sev ol nl slolvlslo ovo .l .lo nolsvlsdvvo

.mbsbvd sl $\mathbf{g}_{\gamma}$ lo $\mathbf{g}_{\rho}$ srbvldvs

$$\hat{\mathbf{u}}_{\mathbf{g}} = \int(\mathrm{sd})\mathbf{v}_{\mathbf{g}}\rho \cdots \mathrm{l}\ \hat{\mathbf{u}}_{\mathbf{g}}(\mathrm{l-q})$$

lo sssvlldsn sn mlsvkjo sl mslvo nl $\hat{\mathbf{v}}_{s,a}$ nsblmlvbnl.sl sl .mvsvlvlo nsdlo vsln

.vld sl mslly $\hat{\mathbf{v}}_{s,a}$ $\cdots$ nl .ol sv.mslvlvs bsdd slvmvlvo lvo .svsvlmv sdl

.lmlvo sld nl svvo sl dsvnqlp vsdvsl sldl .b $\hat{\mathbf{u}}_{\mathbf{g}}$ $\hat{\mathbf{v}}_{s,a}$ lo slvlmvlv bsvslmlvo

sl mslvbsdlsv bslsdldlgvo os sl lsol lldvvo vjsssdd lvnd ,slblmlvbnl sdl lo

bsdlmlsd os $\hat{\mathbf{u}}_{\mathbf{g}}$ vsbsvsd .slvslvmvlo ylvsvldld sld vvl lsv slv .slvslvmvlo

s vlvdo vsslvnl $\hat{\mathbf{v}}_{s-1,s}$ sl sldvmvl .l slvslvmvlv slpssv sdd sd of (d) lvd

sdlmldsol vvd slmvsvdn sld svv lvs .(nvdo lsslvlvo s.d lo svlvdvnd bsrvslldvvo

vslln $\hat{\mathbf{v}}_{s,s}$ lo .slvslvmv sld svvlvlvo sl vsslvlso sl vsslvlso slgvss s lo svslvlvv

.sd vlvlvlv sbvsl

$$K_{n,s} = \text{est. var}(V_{n,s}) = \frac{\Sigma(P_{n,s;-j} - T^*_{n,s})^+}{g(g-1)(g-2)^2} - \frac{gV^2_{n,s}}{(g-2)^2} \quad . \qquad (13)$$

From the contemplated $\sigma^2 \chi^2_d \doteq d$ approximation to $gV_{n,s}$ we derive as the estimate for d

$$d_{n,s} = 2V^2_{n,s} \doteq K_{n,s} \quad . \qquad (14)$$

It should be emphasized that this is a rather crude use of double-jackknifing, and that superior t-approximations may be available. However, $d_{n,s}$ is very easy to compute, and works reasonably well for some situations, as we shall see in the next section. One thing is reasonably clear, that estimation of the appropriate d is likely to be unstable for small values of g. Also the chi-square approximation to $gV_{n,s}$ is likely to be inaccurate in the tails.

4. Some Simulation Results

In order to determine if the foregoing development has some practical substance, we consider some simulated results concerning the sample variance and the sample bivariate correlation. The simulations reported are rather preliminary in scope, and are often based on only 500 repetitions of each situation, but some fairly clear indications are obtained concerning the usefulness of the Student-t confidence limits.

Case 1: The sample variance

Given the set of data $Y_1,\ldots,Y_n$, let

$$S_n^2 = (n-1)^{-1} \Sigma (Y_j - \bar{Y}_\cdot)^2$$

and define $T_n = \log_e S_n^2$, which is the variance-stabilized form of S_n^2. A fairly detailed account of the jackknife in this situation has been given by Miller (1968), although his numerical results are rather suspect. The kurtosis index γ_2 is typically rather large here; for example $\gamma_2 = 6$ for normally distributed data.

The simulations reported here involve three sampling distributions, namely

NOR = normal, DEXP = double exponential, EXP = exponential.

Also three sample sizes were used, $n = 10$, 24 and 50. The number of simulated repetitions of each specific case was usually 1000. Table 1 gives a representative set of empirical means and variances of T_n, $T^*_{n,s}$, and $V_{n,s}$. Also given in each case is the mean of $K_{n,s}$, the estimate of $\text{var}(V_{n,s})$ defined in Section 3. General features to note are (i) the approximate lack of bias of $V_{n,s}$, (ii) the slow increase of $\text{var}(V_{n,s})$ with s, (iii) the overestimation of $\text{var}(V_{n,s})$ by $K_{n,s}$. One might, then, anticipate relatively

4. Some Simulation Results

Additional experience with the Monte Carlo method was gained by applying it to a number of problems for which analytic solutions are known. The results of these calculations are presented in this section.

Example 1: The Gamma Distribution

Let $x_1, x_2, \ldots, x_n$ be a random sample from the gamma distribution

$$g_n(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}$$

The method of maximum likelihood is used to obtain estimates $\hat{\lambda}_n$ of λ. The method is known to be consistent and asymptotically efficient, and tables have been published (Bock) showing the bias and mean square error of $\hat{\lambda}_n$ for various sample sizes n. The Monte Carlo method was used to obtain estimates of the bias and mean square error of $\hat{\lambda}_n$ for $n = 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30$. The number of trials was chosen so that the standard deviation of the estimate of the bias would be approximately .01. Antithetic variates were used to reduce the variance of the estimate of the mean square error. The results are shown in Table 1.

Example 2: The Normal Distribution

Let $x_1, x_2, \ldots, x_n$ be a random sample from the normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The maximum likelihood estimates of μ and σ^2 are

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

The bias of $\hat{\sigma}^2$ is $-\sigma^2/n$. The mean square error of $\hat{\sigma}^2$ is $2(n-1)\sigma^4/n^2$. Monte Carlo estimates of the bias and mean square error of $\hat{\sigma}^2$ were obtained for $n = 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30$. The number of trials was chosen so that the standard deviation of the estimate of the bias would be approximately .01. Antithetic variates were used to reduce the variance of the estimate of the mean square error. The results are shown in Table 2.

small loss of efficiency for jackknife standard errors computed from group pseudo-values ($s > 1$); but confidence limits based on $K_{n,s}$ seem to have a poor prospect.

Table 1 about here

The corresponding results for confidence limits are given in Table 2, the three confidence limit types being denoted

t_∞ = normal approximation

t_{g-1} = Student-t approximation with d.f. = no. groups - 1

t_d = Student-t approximation with d.f. = estimated $d_{n,s}$ of (14).

The results are given in terms of empirical error rates, i.e. non-coverage rates, for nominal error rates 0.2, 0.1, .05 and .01. Standard errors for the empirical rates are roughly equal to error rate $\div$ 30, but for a given sample size all simulated samples are constructed from the same pseudo-random numbers.

The general impression given by Table 2 is that the t_{g-1} approximation is preferable to t_∞, that it tends to be conservative for $s = 1$, but becomes very accurate as s increases. These impressions are particularly strong in the double-exponential situations. Further, the "estimated d.f." approach is very successful when g is large, but is rather unsuccessful when g drops below 8, particularly for low error rates. Other cases not reported here give similar impressions.

Table 2 about here

Теорема 4 (Гаусса).

Если p – простое число, то $(p-1)! \equiv -1 \pmod{p}$.

Доказательство. Пусть a – произвольное целое число, взаимно простое с p. Рассмотрим числа $a, 2a, \ldots, (p-1)a$. Ни одно из них не делится на p, так как p – простое число, а a и k ($k = 1, 2, \ldots, p-1$) взаимно просты. Кроме того, среди этих чисел нет двух сравнимых по модулю p. Действительно, если $ma \equiv na \pmod{p}$, где $1 \le m < n \le p-1$, то $(n-m)a \equiv 0 \pmod{p}$, что невозможно, так как $n - m < p$ и a взаимно просто с p.

Следовательно, числа $a, 2a, \ldots, (p-1)a$ дают при делении на p все ненулевые остатки $1, 2, \ldots, p-1$ в некотором порядке. Поэтому произведение этих чисел сравнимо с произведением $1 \cdot 2 \cdot \ldots \cdot (p-1)$ по модулю p:

$$a \cdot 2a \cdot \ldots \cdot (p-1)a \equiv 1 \cdot 2 \cdot \ldots \cdot (p-1) \pmod{p},$$

или

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}.$$

Так как $(p-1)!$ не делится на p, то можно сократить обе части сравнения на $(p-1)!$, что дает $a^{p-1} \equiv 1 \pmod{p}$.

Теорема доказана.

Теорема 5 (Эйлера).

Если a и m – взаимно простые целые числа, $m > 0$, то $a^{\varphi(m)} \equiv 1 \pmod{m}$, где $\varphi(m)$ – функция Эйлера.

Доказательство. Пусть $r_1, r_2, \ldots, r_{\varphi(m)}$ – приведенная система вычетов по модулю m. Тогда числа $ar_1, ar_2, \ldots, ar_{\varphi(m)}$ также образуют приведенную систему вычетов по модулю m. Поэтому

$$ar_1 \cdot ar_2 \cdot \ldots \cdot ar_{\varphi(m)} \equiv r_1 \cdot r_2 \cdot \ldots \cdot r_{\varphi(m)} \pmod{m},$$

или

$$a^{\varphi(m)} r_1 r_2 \ldots r_{\varphi(m)} \equiv r_1 r_2 \ldots r_{\varphi(m)} \pmod{m}.$$

Сокращая на $r_1 r_2 \ldots r_{\varphi(m)}$, получаем $a^{\varphi(m)} \equiv 1 \pmod{m}$.

Теорема доказана.

Теорема 6.

Пусть $f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0$ – многочлен с целыми коэффициентами, $a_n \neq 0$. Если $x \equiv x_0 \pmod{m}$, то $f(x) \equiv f(x_0) \pmod{m}$.

Доказательство. Из $x \equiv x_0 \pmod{m}$ следует $x^k \equiv x_0^k \pmod{m}$ для любого k. Умножая обе части последнего сравнения на a_k, получаем $a_k x^k \equiv a_k x_0^k \pmod{m}$. Складывая почленно сравнения $a_k x^k \equiv a_k x_0^k \pmod{m}$ для $k = 0, 1, \ldots, n$, получаем $f(x) \equiv f(x_0) \pmod{m}$.

Теорема доказана.

Теорема 7.

Если $f(x) \equiv 0 \pmod{m}$ и $x \equiv x_0 \pmod{m}$, то $f(x_0) \equiv 0 \pmod{m}$.

Доказательство. Из $f(x) \equiv 0 \pmod{m}$ и $f(x) \equiv f(x_0) \pmod{m}$ по свойству транзитивности сравнений получаем $f(x_0) \equiv 0 \pmod{m}$.

Теорема доказана.

Теорема 8.

Если $f(x) \equiv 0 \pmod{m}$ и $x \equiv x_0 \pmod{m}$, то $f(x_0) \equiv 0 \pmod{m}$.

Доказательство. Из $f(x) \equiv 0 \pmod{m}$ и $f(x) \equiv f(x_0) \pmod{m}$ по свойству транзитивности сравнений получаем $f(x_0) \equiv 0 \pmod{m}$.

Теорема доказана.

Confidence limits based on estimated degrees of freedom d are more variable in length than the others, since $d_{n,s}$ has positive correlation with $V_{n,s}$. A fairly natural comparative measure is $t_d \div t_{g-1}$, the ratio of lengths of the the two Student-t intervals. However this ratio needs an adjustment for the different error rates attained by the two methods. We have plotted average t_d and t_{g-1} against empirical error rates to determine if the former is systematically larger when adjusted to a fixed error rate. Generally speaking when $g \geq 10$ the excess is at most 4% for error rates greater than .10, up to 10% for .05 error rate, but as high as 40% when the error rate is .01. In this latter situation the variability of t_d is high. As a typical example, for the same double-exponential samples of size $n = 24$, with $s = 1$, we obtain the comparisons in Table 3; adjusted values of $t_d \div t_{g-1}$ were obtained by crude graphical interpolation to the nominal error rate.

Table 3 about here

The price paid for near-achievement of the nominal error rate is apparently rather slight until we approach the .01 level. One rather curious feature noted in the full set of simulations is that the t_d values seem to depend very little on s; for the case summarized in Table 3 the average t_d values are virtually identical at $s = 3$ and 6.

Case 2: The sample correlation

For n pairs (Y_{1j}, Y_{2j}) the sample correlation coefficient is

$$R_n = \frac{\Sigma(Y_{1j} - \bar{Y}_{1.})(Y_{2j} - \bar{Y}_{2.})}{\{\Sigma(Y_{1j} - \bar{Y}_{1.})^2 \Sigma(Y_{2k} - \bar{Y}_{2.})^2\}^{\frac{1}{2}}} ,$$

$$\mu = \frac{(x_{i+1} - x_i)(x_{j+1} - x_j)}{\left[(x_{i+1} - x_i)(x_{j+1} - x_j)\right]^2}$$

Si μ tend vers 0 lorsque (x_{i+1}, x_{j+1}) prend successivement toutes les valeurs possibles dans Ω, on dira que μ est absolument continue par rapport à ν.

Si μ est absolument continue par rapport à ν, on peut montrer qu'il existe une fonction f définie sur Ω, telle que $\mu(A) = \int_A f \, d\nu$ pour tout ensemble mesurable A de Ω. Cette fonction f est appelée la dérivée de Radon-Nikodym de μ par rapport à ν, et on écrit $f = \frac{d\mu}{d\nu}$.

Dans le cas qui nous intéresse, μ est la mesure de probabilité définie par la fonction de répartition F, et ν est la mesure de Lebesgue. Si F est absolument continue, sa dérivée de Radon-Nikodym par rapport à la mesure de Lebesgue est précisément la densité de probabilité f :

$$f = \frac{dF}{dx}$$

Cette densité f est définie presque partout sur $\mathbb{R}$.

and we take T_n to be $\tanh^{-1}R_n$. Not only does T_n have stable variance for eliptically symmetric distributions, but to second order it is symmetrically distributed. This underlies good performance by the jackknife method. A detailed assessment of the jackknifed correlation has been given by Duncan and Layard (1973), who report simulation studies of the approximate normal (t_∞) confidence intervals.

The simulations reported here all involve elliptically symmetric distributions, for which the conditional expectations are linear. The three distributions used are, with $\rho = \text{corr}(Y_1,Y_2)$ in every case,

NORMAL - standard bivariate normal

CONTAMINATED NORMAL - (Z_1,Z_2) standard bivariate normal and $(Y_1,Y_2) = (Z_1,Z_2)$ with probability 0.9, $(Y_1,Y_2) = 3(Z_1,Z_2)$ with probability 0.1

DOUBLE EXPONENTIAL - Y_1 double exponential, $Y_2 = (1-\rho^2)^{\frac{1}{2}} Y_1 + \rho Z$ where Z is double exponential independent of Y_1.

Samples sizes used were $n = 10$, 24 and 50, and correlation values ranged from 0.0 to 0.9. We report here simulations for $\rho = 0.6$, with additional results for $\rho = 0.9$ at $n = 10$. The behavior of the jackknife is fairly uniform with respect to ρ until ρ exceeds 0.9; see also Duncan and Layard (1973). All results for $n = 10$, 24 are based on 500 simulations, while those for $n = 50$ are based on 250 simulations. It is hard to make precise conclusions from such a small-scale experiment, but a fairly clear picture does emerge.

and we shall be concerned with the case when μ is not necessarily absolutely continuous with respect to μ^*. Note that μ^* is not unique, but it is unique if μ is absolutely continuous with respect to μ^* and μ^* is absolutely continuous with respect to μ.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *singular* if there exists a Borel set S such that $\mu(S) = 1$ and $\lambda_k(S) = 0$, where λ_k denotes Lebesgue measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Two probability measures μ and ν on $(\mathbb{R}^k, \mathcal{B}^k)$ are said to be *mutually singular* if there exist disjoint Borel sets S and T such that $\mu(S) = 1$ and $\nu(T) = 1$. In this case we write $\mu \perp \nu$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *absolutely continuous* with respect to Lebesgue measure if for every Borel set B with $\lambda_k(B) = 0$, we have $\mu(B) = 0$. In this case we write $\mu \ll \lambda_k$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *discrete* if there exists a countable set S such that $\mu(S) = 1$. In this case we write $\mu \ll \nu$ where ν is counting measure on S.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *continuous* if $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}^k$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *absolutely continuous with respect to Lebesgue measure* if there exists a Borel measurable function f such that $f \geq 0$ and $\int_{\mathbb{R}^k} f \, d\lambda_k = 1$, and $\mu(B) = \int_B f \, d\lambda_k$ for every Borel set B. The function f is called a *density* of μ with respect to λ_k.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *singular with respect to Lebesgue measure* if $\mu \perp \lambda_k$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *absolutely continuous with respect to counting measure* if $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}^k$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *singular with respect to counting measure* if $\mu \perp \nu$ where ν is counting measure on S.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *absolutely continuous with respect to counting measure* if $\mu \ll \nu$ where ν is counting measure on S.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *continuous* if $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}^k$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *absolutely continuous with respect to Lebesgue measure* if there exists a Borel measurable function f such that $f \geq 0$ and $\int_{\mathbb{R}^k} f \, d\lambda_k = 1$, and $\mu(B) = \int_B f \, d\lambda_k$ for every Borel set B. The function f is called a *density* of μ with respect to λ_k.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *singular with respect to Lebesgue measure* if $\mu \perp \lambda_k$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *absolutely continuous with respect to counting measure* if $\mu(\{x\}) = 0$ for every $x \in \mathbb{R}^k$.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *singular with respect to counting measure* if $\mu \perp \nu$ where ν is counting measure on S.

Let μ be a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$. Then μ is said to be *absolutely continuous with respect to counting measure* if $\mu \ll \nu$ where ν is counting measure on S.

Table 4 displays empirical means and variances of T_n, $T^*_{n,s}$ and $V_{n,s}$ for several representative cases. General features to notice are (i) the bias-removal effect of $T^*_{n,s}$; (ii) the tendency for $\text{var}(T^*_{n,s})$ to exceed $\text{var}(T_n)$; (iii) the general lack of bias in $V_{n,s}$ (there are few exceptions); (iv) the slow increase of $\text{var}(V_{n,s})$ with increasing s, except in the normal case, where $\gamma_2 = 2$; (v) the overestimation of $\text{var}(V_{n,s})$ by $K_{n,s}$.

Rather oddly, when the direct sample second and fourth moments of pseudo-values are used to estimate $\text{var}(V_{n,s})$, as described by Rao (1975, p.438) for example, the estimate is nearly unbiased. This suggests that the approximate theory of Section 2 is quite accurate, even for n as small as 10. However it turns out that $K_{n,s}$ gives slightly better performance in connection with confidence limits.

Table 4 about here

For the same simulations as above Table 5 contains empirical error rates for the three types of confidence intervals, denoted t_f with degrees of freedom $f = \infty$ (normal approximation), $g-1$ (no. groups minus one) and $d = d_{n,s}$ defined by (14). The general impression is that the t_{g-1} approximation is superior to t_∞, and is the more accurate as s increases. Also the estimated degrees of freedom method works very well for $g \geq 10$.

Table 5 about here

As with the case of the sample variance earlier, the values of t_d are quite variable, particularly for error rates below .05. Since the error rates for t_{g-1} and t_d differ, the corresponding mean interval lengths are

Table 4 displays empirical means and variances of T_n, $T_{n,s}$ and $V_{n,s}$ for several representative cases. General features to notice are (i) the bias-removal effect of $T_{n,s}$; (ii) the tendency for $var(T_{n,s})$ to exceed $var(T_n)$; (iii) the general lack of bias in $V_{n,s}$ (there are few exceptions); (iv) the slow increase of $var(V_{n,s})$ with increasing s, except in the normal case, where $\gamma_2 = 3$; (v) the overestimation of $var(V_{n,s})$ by $K_{n,s}$.

Rather oddly, when the direct sample second and fourth moments of pseudo-values are used to estimate $var(V_{n,s})$, as described by Lee (1975, p.438) for example, the estimate is nearly unbiased. This suggests that the approximate theory of Section 2 is quite accurate, even for n as small as 10. However it turns out that $K_{n,s}$ gives slightly better performance in connection with confidence limits.

Table 4 about here

For the same simulations as above Table 5 contains empirical error rates for the three types of confidence intervals, denoted t_d with degrees of freedom $d = \infty$ (normal approximation), $g-1$ (no. groups minus one) and $d = d_{n,s}$ defined by (14). The general impression is that the t_{g-1} approximation is superior to t_∞ and is the more accurate as s increases. Also the estimated degrees of freedom method works very well for $g \geq 10$.

Table 5 about here

As with the case of the sample variance earlier, the values of t_d are quite variable, particularly for error rates below .05. Since the error rates for t_{g-1} and t_d differ, the corresponding mean interval lengths are

not directly comparable. As before, we have adjusted the mean lengths to common error rate by simple graphical interpolation. Some summaries for adjusted mean ratios of lengths are given in Table 6; results for other cases are broadly similar. On the average a small price is paid for achieving nominal error rates of .05 and higher; for error rate .01 very long intervals may be obtained using t_d.

Table 6 about here

---

mature or magnetic mean and bacteriba arrive at ,erotsh .aldatsopsidomrofle compatible. As before, we have selected the mean radiance to current bacteriba not be satisfactory fundamentalogration. Some examples for open cases are seen below. Although the model is clear at first glance it should be noted that the mean radiance of samples that have a significant amount of clay tend to have lower values for open cases. Although quartz elastasle that have very long fundamental periods of 10. Even for rocks that have 50. No gaines beneficio on

Apple o spond pend

Apple o spond pend

oxed dnods o alda7

5. Tentative Conclusions and Further Remarks

The crude theoretical development in Section 2 reliably predicts two important practical results: the approximate normal jackknife confidence limits are often inaccurate even for moderately large n, and the Student-t approximation with $g-1$ degrees of freedom is better for larger s. An exception is the case of the bivariate transformed correlation when data has a bivariate normal distribution, for which the very low value of γ_2 indicates a very accurate normal approximation. Also the efficiency of the jackknife standard error $\sqrt{V_{n,s}}$ may be very high for $s > 1$, as in the case of the sample variance for a sample of $n = 24$ double-exponential variables. Thus grouping of the data usually improves the probability accuracy of confidence limits, and may involve slight loss of efficiency. For example with a sample of $n = 50$ pairs, a jackknife confidence interval for correlation ρ via $z = \tanh^{-1} r$ using ten groups of size 5 would be reasonably efficient and robust in terms of true coverage probability.

Estimation of the degrees of freedom for Student-t intervals can give good results for more than ten groups, apparently. Often the resulting procedure is somewhat conservative and is seemingly unstable for error rates smaller than .05. A similar method of obtaining prediction confidence intervals has been considered by Butler and Rothman (1975) in their work on cross-validation.

These tentative conclusions rest on our rather sketchy simulation experiments. Clcearly more intensive simulation is needed before precise recommendations can be made.

Although the use of group size greater than one has reliability advantages, it has two disadvantages. First, <u>random</u> grouping of the data is

5. Tentative Conclusions and Further Remarks

The crude theoretical development in Section 2 reliably predicts two important practical results: the approximate normal jackknife confidence limits are often inaccurate even for moderately large n, and the student-t approximation with $g-1$ degrees of freedom is better for larger s. An exception is the case of the bivariate transformed correlation when data has a bivariate normal distribution, for which the very low value of γ_2 indicates a very accurate normal approximation. Also the efficiency of the jackknife standard error $\hat{\sigma}_{n,s}$ may be very high for $s > 1$, as in the case of the sample variance for a sample of $n = 24$ double-exponential variables. Thus grouping of the data usually improves the probability accuracy of confidence limits, and may involve slight loss of efficiency. For example with a sample of $n = 50$ pairs, a jackknife confidence interval for correlation ρ via $z = \tanh^{-1} r$ using ten groups of size 5 would be reasonably efficient and robust in terms of true coverage probability.

Estimation of the degrees of freedom for Student-t intervals can give good results for more than ten groups, apparently. When the resulting procedure is somewhat conservative and is seemingly unstable for error rates smaller than .05. A similar method of obtaining prediction confidence intervals has been considered by Butler and Rothman (1975) in their work on cross-validation. These tentative conclusions rest on our rather slim simulation experiments. Clearly more intensive simulation is needed before precise recommendations can be made. Although the use of group size greater than one has reliability advantages, it has two disadvantages. First, random grouping of the data is

involved, which is objectionable on principle: the resulting inference is not unique. Secondly, pseudo-values for grouped data are not as capable of indicating deviant points. It is useful to realize that the pseudo-value residual $P_{n;j} - T_n^*$ measures the influence of Y_j on T_n, and so may be used in identification of outliers; see Devlin, Gnanadesikan and Kettenring (1975).

Somewhat connected to the last remark is the idea that pseudo-values may be used to define robust estimates analogous to those for "exact" location parameters. This possibility depends on the detailed structure of representation (12). Work in progress indicates that robust correlation estimates can be obtained from robust combination of pseudo-values.

This research was supported by National Science Foundation Grant No.MPS75-08778. Programming for the calculations was done by Thomas A. Davis, under the auspices of a University of Minnesota Graduate School Research Assistantship.

involved, which is objectionable on principle: the resulting inference is not unique. Secondly, pseudo-values for grouped data are not as capable of indicating deviant points. It is useful to realize that the pseudo-value residual $R_{n;j} - T_n^*$ measures the influence of Y_j on T_n, and so may be used in identification of outliers; see Devlin, Gnanadesikan and Kettenring (1975).

Somewhat connected to the last remark is the idea that pseudo-values may be used to define robust estimates analogous to those for "exact" location parameters. This possibility depends on the detailed structure of representation (12). Work in progress indicates that robust correlation estimates can be obtained from robust combination of pseudo-values.

This research was supported by National Science Foundation Grant No. [illegible]. Programming for the calculations was done by Thomas A. Davis, under the auspices of a University of Minnesota Graduate School Research Assistantship.

References

Butler, R. and Rothman, E.D. (1975) "Prediction intervals based on reuse of the sample." University of Michigan, Department of Statistics, Tech. Rep. No. 56.

Devlin, S.J., Gnanadesikan, R. and Kettenring, J.R. (1975) "Robust estimation and outlier detection with correlation coefficients." Biometrika 62, 531-545.

Duncan, J.T. and Layard, M.W.J. (1973) "A Monte Carlo study of asymptotically robust tests for correlation coefficients." Biometrika 60, 551-558.

Kish, L. and Frankel, M.R. (1974) "Inference from complex samples (with duscussion)." J. Roy. Statist. Soc. B 36, 1-37.

Miller, R.G. (1974) "The jackknife - a review." Biometrika 61, 1-15.

Miller, R.G. (1968) "Jackknifing variances." Ann. Math. Statist. 39, 567-582.

Mosteller, F. and Tukey, J.W. (1964) "Data analysis, including statistics." in Handbook of Social Psychology (O. Lindzey and E. Aronson, eds.), vol. 2. Reading, Mass: Addison-Wesley.

Quenouille, M.H. (1956) "Notes on bias in estimation." Biometrika 43, 353-360.

Rao, C.R. (1975) Linear Statistical Inference (second edition). New York: Wiley.

Rao, J.N.K. and Webster, J. (1966) "On two methods of bias reduction in the estimation of ratios." Biometrika 53, 571-577.

Tukey, J.W. (1958) "Bias and confidence in not-quite large samples (abstract)." Ann. Math. Statist. 29, 614.

Literatur

Bauer, H. und Schaffernicht, J. (1954): Über die Reaktion von Kohlendioxid mit Alkalihydroxiden in wäßriger Lösung. Z. Naturforsch. 9a, S. 399-408.

Bergmann, E. und Polanyi, M. (1933): Über Reaktionsgeschwindigkeiten in Lösungen. Z. Phys. Chem. B 23, S. 21-41.

Bodenstein, M. (1913): Eine Theorie der photochemischen Reaktionsgeschwindigkeiten. Z. Phys. Chem. 85, S. 329-397.

Brönsted, J.N. und Volqvartz, K. (1934): Über die Geschwindigkeit sauer katalysierter Reaktionen in wäßriger Lösung. Z. Phys. Chem. A 169, S. 1-31.

Eigen, M. (1954): Über die Geschwindigkeit von Neutralisationsvorgängen. Z. Phys. Chem. N.F. 1, S. 176-200.

Eigen, M. (1963): Protonenübertragungen, Säure-Base-Katalyse und enzymatische Hydrolyse. Teil I: Elementarprozesse. Angew. Chem. 75, S. 489-508.

Eigen, M. und Kustin, K. (1962): The Kinetics of Halogen Hydrolysis. J. Am. Chem. Soc. 84, S. 1355-1361.

Eigen, M. und Maeyer, L. de (1955): Self-dissociation and Protolytic Reactions of Water and Heavy Water in Liquid and Solid States. Proc. Roy. Soc. A 232, S. 176-187.

Eigen, M. und Maeyer, L. de (1958): Relaxation Methods. In: Weissberger, A. (Ed.): Technique of Organic Chemistry, Vol. VIII, Part II: Investigation of Rates and Mechanisms of Reactions. 2nd Ed. New York: Interscience Publ. S. 895-1054.

Eigen, M. und Tamm, K. (1962): Schallabsorption in Elektrolytlösungen in Abhängigkeit von Konzentration und Molekularprozessen. Z. Elektrochem. 66, S. 93-121 und S. 107-121.

Eigen, M. und Wilkins, R.G. (1965): The Kinetics of Complex Formation and Exchange Reactions of Metal Ions: Ligand Effects. In: Kustin, K. (Ed.): Mechanisms of Inorganic Reactions. Adv. Chem. Ser. 49, S. 55-80.

Fabuss, B.M., Smith, J.O. und Satterfield, C.N. (1964): Proton Magnetic Resonance Spectra of Sugars in Deuterium Oxide. Adv. Chem. Ser. 74, S. 1-20.

Fabuss, B.M., Smith, J.O. und Satterfield, C.N. (1964): Proton Magnetic Resonance Spectra of Sugars in Deuterium Oxide. Adv. Chem. Ser. 74, S. 1-20.

Table 1. Empirical Properties of $T^*_{n,s}$ and $V_{n,s}$ when $T_n = \log_e S^2_n$ with normal, double-exponential and exponential data; 1000 simulations in each case

case	group size s	$E(T_n)$ emp.	$E(T_n)$ theor.	$E(T^*_{n,s})$ emp.	$var(T_n)$ emp.	$var(T^*_{n,s})$ emp.	$E(V_{n,s})$ emp.	$var(V_{n,s})$ emp.	$var(V_{n,s})$ est.*
NOR	1	0	0	.02	.041	.041	.042	.00022	.00048
n=50	5			.02		.041	.042	.00057	.00090
NOR	1	-.011	0	.036	.081	.080	.099	.0032	.0063
n=24	3			.035		.081	.099	.0053	.0090
	6			.033		.083	.097	.0091	.0145
DEXP	1	.650	.693	.699	.094	.102	.106	.0087	
n=50	5			.699		.104	.107	.0130	
DEXP	1	.595	.693	.692	.204	.233	.224	.052	.084
n=24	3			.693		.238	.224	.068	.102
	6			.693		.240	.226	.083	.124
DEXP	1	.416	.693	.640	.536	.644	.576	.404	.472
n=10									
EXP	1	-.141	0	-.008	.311	.396	.336	.224	.319
n=24	3			-.008		.411	.336	.296	.395
	6			-.008		.416	.333	.291	.394

*empirical mean of $K_{n,s}$

Table 1. Asymptotic Properties of $\hat{\mu}_{n,s}^{n}$ and $V_{n,s}$ when $Y_n^n = \log_e s_n^2$ with normal, double-exponential and uniform(-.5,.5) distributions in each case

case	group size	$E(\hat{\mu}_{n,s}^{n})$ emp.	$E(\hat{\mu}_{n,s}^{n})$ theor.	$E(Y_{n,s}^{n})$ emp.	$\text{var}(Y_{n,s}^{n})$ emp.	$C(\hat{\mu}_{n,s}^{n}, Y_{n,s}^{n})$ emp.	$\text{var}(\hat{\mu}_{n,s}^{n})$ emp.	$\text{var}(\hat{\mu}_{n,s}^{n})$ asymp.
NOR	1	0	0	.02	.041	.041	.042	.00022
n=50	5			.02	.041		.042	.00037
NOR	1	-.011	0	.089	.000	.001	.089	.0232
n=24	3			.085	.013		.089	.0232
	6			.083	.083		.087	.0097
DBLEX	1	.650	.693	.589	.102	.094	.106	.0087
n=50	5			.579	.104		.107	.0150
DBLEX	1	.353	.693	.502	.233	.204	.224	.052
n=24	3			.699	.238		.224	.090
	6			.703	.240		.225	.022
UNIF	1	.416	.655	.640	.534	.044	.376	.404
n=50								
UNIF	1	-.161	0	-.008	.311	.395	.329	.224
n=24	3			-.008	.411	.396	.207	.395
	6			-.008	.411	.393	.201	.396

*Empirical mean of $K_{n,s}$

Table 2. Empirical error rates for jackknife confidence intervals for variance

sample size group size	nominal error rate	normal			double-exponential			exponential		
		t_∞	t_{g-1}	t_d	t_∞	t_{g-1}	t_d	t_∞	t_{g-1}	t_d
	.2	.220	.177	.159	.280	.254	.237	.280	.254	.214
n=10	.1	.116	.096	.085	.187	.156	.145	.187	.156	.129
s=1	.05	.082	.056	.050	.140	.105	.098	.140	.105	.087
g=10	.01	.034	.012	.011	.091	.045	.046	.081	.045	.037
	.2	.181	.168	.150	.255	.242	.204	.320	.319	.239
n=24	.1	.098	.087	.076	.166	.148	.112	.217	.197	.146
s=1	.05	.057	.050	.040	.108	.092	.064	.157	.146	.093
g=24	.01	.015	.011	.008	.045	.031	.018	.092	.081	.036
	.2	.211	.165	.166	.295	.256	.234	.333	.291	.258
n=24	.1	.116	.072	.080	.197	.155	.150	.239	.193	.173
s=3	.05	.065	.037	.039	.139	.085	.090	.181	.133	.109
g=8	.01	.027	.003	.006	.067	.020	.031	.113	.050	.051
	.2	.271	.174	.186	.330	.237	.243	.365	.269	.269
n=24	.1	.173	.076	.087	.234	.129	.138	.268	.163	.179
s=6	.05	.120	.039	.046	.182	.063	.081	.210	.090	.121
g=4	.01	.066	.005	.016	.098	.012	.027	.143	.022	.049
	.2	.228	.224	.205	.242	.234	.200			
n=50	.1	.115	.104	.089	.134	.128	.106			
s=1	.05	.061	.056	.043	.092	.084	.062			
g=50	.01	.030	.012	.016	.054	.020	.038	(not sampled)		
	.2	.242	.215	.210	.268	.242	.238			
n=50	.1	.141	.092	.104	.150	.120	.126			
s=5	.05	.077	.049	.048	.100	.072	.084			
g=10	.01	.030	.012	.016	.054	.020	.038			

(Y)Dsw nod Alrrtrtrnd aouuhlzpo ohlnnbhdnl nob nnqbr zohrno Isnhlhqqnn .n nlcsn

| | | |

Table 3. Comparison of lengths of Student-t intervals
Double-exponential samples size n=24, s=1

normal error rate	.2	.1	.05	.01
t_{g-1}	1.319	1.714	2.069	2.807
ave. t_d	1.474	2.021	2.599	4.236
s.e. t_d	0.14	0.30	0.54	1.66
mean adjusted $t_d \div t_{g-1}$	1.00	1.03	1.11	1.37
s.e. adjusted $t_d \div t_{g-1}$	0.11	0.17	0.26	0.59

Analysis of Covariance Model Incorporating Heteroscedasticity and Autocorrelation

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4$ (equality of slopes)

Source of variation	Sum of squares	d.f.	Mean square	Variance ratio
Between slopes	1.3145	3	0.4382	0.4382/0.0360 = 12.17
Between regressions	0.0360	1	0.0360	
Deviation from slopes	1.2785	3	0.4262	
Within groups	0.0360	1	0.0360	
Total	1.3145	4		

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$ (common slope)

Source of variation	Sum of squares	d.f.	Mean square	Variance ratio
Between slopes	1.3145	3	0.4382	0.4382/0.0360 = 12.17
Between regressions	0.0360	1	0.0360	
Deviation from slopes	1.2785	3	0.4262	
Within groups	0.0360	1	0.0360	
Total	1.3145	4		

Table 4. Empirical Properties of $T^*_{n,s}$ and $V_{n,s}$ when $T_n = \tanh^{-1}(R_n)$ with bivariate normal (NOR), contaminated bivariate normal (CONT) and bivariate double exponential (DEXP); 500 simulations for n = 10, 24 and 250 simulations for n = 50.

case	group size s	$E(T_n)$ emp.	$E(T_n)$ theor.	$E(T_{n,s})$ emp.	$var(T_n)$ emp.	$var(T_{n,s})$ emp.	$E(V_{n,s})$ emp.	$var(V_{n,s})$ emp.	$var(V_{n,s})$ est.*
NOR n=10 r=.9	1	1.527	1.472	1.439	.1338	.1390	.1746	.0152	.0257
NOR n=24 r=.6	1	0.706	0.693	0.689	.0487	.0472	.0516	.00053	.00097
	3			0.691		.0479	.0534	.00117	.00177
NOR n=50 r=.6	1	0.703	0.693	0.696	.0216	.0212	.0218	.000034	.000075
	5			0.696		.0213	.0225	.000122	.000167
CONT n=10 r=.9	1	1.589	1.472	1.498	.219	.319	.282	.086	.124
CONT n=24 r=.6	1	0.726	0.693	0.703	.108	.137	.109	.0117	.0166
	3			0.705		.137	.106	.0125	.0175
CONT n=50	1	0.711	0.693	0.696	.0589	.0698	.0554	.00120	.0031
	5			0.696		.0701	.0545	.00215	.0033
DEXP n=10 r=.9	1	1.542	1.472	1.457	.238	.287	.288	.0711	.105
DEXP n=24 r=.6	1	0.728	0.693	0.702	.0626	.0689	.0803	.0038	.0056
	3			.0702		.0704	.0795	.0043	.0064
DEXP n=50 r=.6	1	0.707	0.693	0.694	.0305	.0317	.0333	.00035	.00051
	5			0.694		.0322	.0328	.00061	.00082

*empirical mean of $K_{n,s}$

Table 4. Antithetical Bivariate Sub $\bar{X}_{n,s}$ and $Y_{n,s}$ when Σ_{xy} = Zero (ρ_{xy} = Zero) with Bivariate Normal (N_2), Contaminated Bivariate Normal (CN_2) and Bivariate Double Exponential (DE 2): 500 Replications for n = 10, 24 and 250 Simultaneously for n = 50.

case	σ_{xy}	$\sigma_{xy}/\sigma_x\sigma_y$	$B(\bar{X}_{n,s})$ comp.	$B(\bar{X}_{n,s})$ theor.	$Var(\bar{X}_{n,s})$ emp.	$n\,Var(\bar{X}_{n,s})$ emp.	$Var(\bar{Y}_{n,s})$ emp.	$n\,Var(\bar{Y}_{n,s})$ emp.
N(0, Σ) n=10 r=.5	1	1.827	1.672	1.439	.1390	.1760	.0132	.0527
N(0, Σ) n=24 r=.5	1	0.706	0.595	0.595	.0427	.0712	.0052	.0027
	5			0.591		.0470	.0234	.0017
N(0, Σ) n=50 r=.5	1	0.712	0.693	0.596	.0210	.0718	.0025	.0007
	5			.0596		.0213	.0013	.0000
CONTC n=10 r=.5	1	1.389	1.472	1.4021	.211	.138	.036	.124
CONTC n=24 r=.5	1	0.726	0.593	0.705	.106	.127	.109	.0127
	5			0.705		.127	.1100	.0125
CONTC n=50 r=.5	1	0.711	0.695	0.696	.0569	.0569	.0579	.0120
	5			.0696		.0711	.0245	.0013
DE2 n=10 r=.5	1	1.562	1.472	1.457	.292	.307	.588	.0711	.105
DE2 n=24 r=.5	1	0.721	0.595	0.702	.0570	.0702	.0203	.0038	.0024
	5			.0702		.0708	.0795	.0053	.0004
DE2 n=50 r=.5	1	0.707	0.695	0.694	.0305	.0317	.0333	.0015	.0002
	5			.0694		.0322	.0598	.0006	.0000

*Empirical mean of $\bar{X}_{n,s}$

Table 5. Empirical error rates of jackknife confidence intervals for correlation

sample size group size correlation	nominal error rate	normal			cont. normal			dbl.-expl.		
		t_∞	t_{g-1}	t_d	t_∞	t_{g-1}	t_d	t_∞	t_{g-1}	t_d
n=10	.2	.204	.178	.160	.266	.226	.172	.280	.236	.192
s=1	.1	.122	.096	.082	.133	.102	.076	.158	.098	.076
g=10	.05	.076	.048	.042	.086	.060	.036	.084	.054	.042
r=.9	.01	.030	.014	.010	.040	.012	.010	.034	.014	.006
n=24	.2	.200	.188	.174	.306	.284	.224	.212	.188	.154
s=1	.1	.122	.112	.094	.174	.166	.104	.120	.106	.074
g=24	.05	.072	.056	.044	.104	.096	.054	.060	.048	.030
r=.6	.01	.028	.016	.010	.040	.022	.010	.022	.014	.008
n=24	.2	.220	.184	.180	.342	.280	.246	.236	.200	.162
s=3	.1	.142	.100	.094	.200	.156	.122	.134	.109	.078
g=8	.05	.086	.052	.052	.144	.088	.056	.084	.044	.042
r=.6	.01	.036	.006	.020	.062	.018	.016	.036	.006	.010
n=50	.2	.212	.212	.208	.300	.284	.232	.232	.224	.200
s=1	.1	.108	.100	.092	.172	.168	.096	.116	.104	.080
g=50	.05	.060	.052	.036	.120	.112	.044	.044	.032	.020
r=.6	.01	.008	.008	.004	.036	.020	.012	.016	.012	.008
n=50	.2	.244	.208	.208	.312	.280	.256	.264	.232	.208
s=5	.1	.144	.128	.128	.216	.168	.116	.136	.092	.084
g=10	.05	.104	.064	.060	.136	.076	.044	.080	.052	.040
r=.6	.01	.044	.008	.004	.044	.004	.008	.036	.008	.008

Table 5. Numerical error rates of jackknife confidence intervals for correlation

correlation group size	error rate	normal			cont. normal			bi.-expl.		
		$\hat{\rho}$	$\hat{\rho}-\rho_1$	$\hat{\rho}-\rho_2$	$\hat{\rho}$	$\hat{\rho}-\rho_1$	$\hat{\rho}-\rho_2$	$\hat{\rho}$	$\hat{\rho}-\rho_1$	$\hat{\rho}-\rho_2$
n=10	.2	.265	.297	.197	.256	.226	.172	.280	.230	.144
s=1	.1	.122	.056	.032	.123	.102	.074	.159	.107	.078
ρ=.5	.05	.075	.042	.042	.079	.060	.056	.094	.054	.042
r=.9	.01	.030	.014	.010	.040	.017	.010	.084	.016	.008
n=24	.2	.200	.109	.074	.204	.204	.224	.224	.180	.134
s=1	.1	.122	.112	.096	.176	.136	.120	.160	.104	.074
ρ=.5	.05	.072	.056	.044	.104	.096	.084	.096	.068	.050
r=.9	.01	.028	.016	.010	.0404	.012	.010	.022	.014	.008
n=24	.2	.220	.134	.090	.242	.160	.096	.236	.200	.102
s=3	.1	.142	.110	.084	.200	.156	.122	.134	.102	.070
ρ=.8	.05	.086	.052	.032	.044	.096	.056	.084	.064	.042
r=.9	.01	.036	.006	.004	.016	.012	.004	.026	.004	.002
n=50	.2	.212	.212	.096	.300	.124	.082	.232	.224	.090
s=1	.1	.108	.100	.062	.172	.116	.076	.116	.106	.062
ρ=.5	.05	.080	.052	.034	.120	.092	.064	.084	.052	.030
r=.9	.01	.008	.008	.004	.026	.020	.012	.018	.012	.006
n=50	.2	.244	.208	.112	.260	.256	.264	.252	.208	
s=3	.1	.144	.128	.076	.168	.128	.096	.092	.054	
ρ=.8	.05	.104	.084	.050	.080	.076	.064	.052	.040	
r=.9	.01	.044	.008	.004	.014	.008	.004	.008	.008	

Table 6. Adjusted mean ratio and s.d. of confidence interval length for t_d and t_{g-1}

	contaminated normal n=24, s=1, ρ=0.6			contaminated normal n=50, s=1, ρ=0.6			double exponential n=10, s=1, ρ=0.9		
nominal error rate =	.1	.05	.01	.1	.05	.01	.1	.05	.01
t_{g-1}	1.714	2.069	2.807	1.677	2.009	2.680	1.833	2.262	3.250
mean t_d	2.065	2.678	4.468	2.021	2.594	4.192	2.536	2.803	4.816
mean adjusted $t_d \div t_{g-1}$	1.00	1.02	1.08	1.00	1.06	1.38	1.08	1.14	1.22
s.d. adjusted $t_d \div t_{g-1}$	.18	.28	.65	.15	.23	.52	.17	.25	.56

Entsprechend der Lage des Punktes P wird der Einfluß
$\frac{z-a}{z-b}$ bzw. $\frac{z-b}{z-a}$ auf dessen Lage betrachtet.

Lage des Punktes P	Einfluß auf die Lage des Punktes P	Lage des Punktes P	Einfluß auf die Lage des Punktes P
$\varphi = 0$	$\varphi = \varphi_1 - \varphi_2$	$\varphi = 0$	$\varphi = \varphi_2 - \varphi_1$
$\varphi_1 = \varphi_2$	$\varphi = 0$	$\varphi_1 = \varphi_2$	$\varphi = 0$
$\varphi_1 > \varphi_2$	$\varphi > 0$	$\varphi_1 > \varphi_2$	$\varphi < 0$
$\varphi_1 < \varphi_2$	$\varphi < 0$	$\varphi_1 < \varphi_2$	$\varphi > 0$

φ = Winkel zwischen beiden Strahlen

$|z-a|$ bzw. $|z-b|$

$|z-a|$	$|z-b|$	$|z-a| \cdot |z-b|$	$\frac{|z-a|}{|z-b|}$	$\frac{|z-b|}{|z-a|}$
10.	10.	100.0	1.00	1.00
10.	20.	200.0	0.50	2.00
10.	30.	300.0	0.33	3.00
10.	40.	400.0	0.25	4.00
10.	50.	500.0	0.20	5.00
20.	20.	400.0	1.00	1.00
20.	30.	600.0	0.67	1.50
20.	40.	800.0	0.50	2.00
20.	50.	1000.0	0.40	2.50
30.	30.	900.0	1.00	1.00
30.	40.	1200.0	0.75	1.33
30.	50.	1500.0	0.60	1.67
40.	40.	1600.0	1.00	1.00
40.	50.	2000.0	0.80	1.25
50.	50.	2500.0	1.00	1.00

$|z-a| \cdot |z-b|$ = Produkt beider Strahlen

$\frac{|z-a|}{|z-b|}$ bzw. $\frac{|z-b|}{|z-a|}$ = Quotient beider Strahlen